AFRL-RX-WP-TP-2008-4040

ESTIMATING FLAW SIZE DISTRIBUTIONS FROM SERVICE INSPECTION RESULTS (PREPRINT)

Peter W. Hovey and Jeremy Knopp

Nondestructive Evaluation Branch
Metals, Ceramics & Nondestructive Evaluation Division

DECEMBER 2006

Approved for public release; distribution unlimited.

See additional restrictions described on inside pages

STINFO COPY

AIR FORCE RESEARCH LABORATORY
MATERIALS AND MANUFACTURING DIRECTORATE
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7750
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YY)*: December 2006

2. REPORT TYPE: Journal Article Preprint

3. DATES COVERED *(From - To)*:

4. TITLE AND SUBTITLE: ESTIMATING FLAW SIZE DISTRIBUTIONS FROM SERVICE INSPECTION RESULTS (PREPRINT)

5a. CONTRACT NUMBER: In-house

5b. GRANT NUMBER:

5c. PROGRAM ELEMENT NUMBER: 62102F

5d. PROJECT NUMBER: 4349

5e. TASK NUMBER: RG

5f. WORK UNIT NUMBER: M04R1000

6. AUTHOR(S)

Peter W. Hovey (University of Dayton)
Jeremy Knopp (AFRL/RXLP)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Dayton
Department of Mathematics
Dayton, OH 45469

Nondestructive Evaluation Branch (AFRL/RXLP)
Metals, Ceramics & Nondestructive Evaluation Division
Materials and Manufacturing Directorate
Wright-Patterson Air Force Base, OH 45433-7750
Air Force Materiel Command, United States Air Force

8. PERFORMING ORGANIZATION REPORT NUMBER: AFRL-RX-WP-TP-2008-4040

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory
Materials and Manufacturing Directorate
Wright-Patterson Air Force Base, OH 45433-7750
Air Force Materiel Command
United States Air Force

10. SPONSORING/MONITORING AGENCY ACRONYM(S): AFRL/RXLP

11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S): AFRL-RX-WP-TP-2008-4040

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

The U.S. Government is joint author of this work and has the right to use, modify, reproduce, release, perform, display, or disclose the work. PAO Case Number: AFRL/WS 07-1065, 01 May 2007.

14. ABSTRACT

A key component of risk analyses of aging aircraft is the distribution of flaw sizes that are present in the aircraft. This distribution can be derived from teardown inspections of retired aircraft; however it is more cost effective to use the results of service inspections. Using the sizes of found cracks can be misleading however because nondestructive inspections are not perfect so some cracks are missed. Furthermore, the likelihood that an individual crack is detected is a function of the size of the crack when inspected and the crack size distribution is related to the number of flight hours the aircraft has experienced. An approach for estimating flaw size distributions from inspection results is derived and illustrated from data and simulation results. Problems with estimating both the POD function and the crack size distribution are discussed and a method for setting the reset crack size after an inspection based on the sizes of detected cracks is suggested.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT:	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON (Monitor)
Unclassified	Unclassified	Unclassified	SAR	16	Kumar Jata

19b. TELEPHONE NUMBER *(Include Area Code)*: N/A

Estimating Flaw Size Distributions from Service Inspection Results

Peter W. Hovey[1] and Jeremy Knopp[2]
Department of Mathematics, University of Dayton, Dayton, OH 45469, USA[1].
Metals, Ceramics, and NDE Division, Air Force Research Laboratory (AFRL/MLLP) [2]

Abstract

A key component of risk analyses of aging aircraft is the distribution of flaw sizes that are present in the aircraft. This distribution can be derived from teardown inspections of retired aircraft; however it is more cost effective to use the results of service inspections. Using the sizes of found cracks can be misleading however because nondestructive inspections are not perfect so some cracks are missed. Furthermore, the likelihood that an individual crack is detected is a function of the size of the crack when inspected and the crack size distribution is related to the number of flight hours the aircraft has experienced. An approach for estimating flaw size distributions from inspection results is derived and illustrated from data and simulation results. Problems with estimating both the POD function and the crack size distribution are discussed and a method for setting the reset crack size after an inspection based on the sizes of detected cracks is suggested.

I. Introduction

The Damage Tolerant Design (DTD) concept specified by the US Air Force for aircraft design and maintenance incorporates redundant load paths and periodic inspections for fatigue cracks to ensure safety. A key part of the periodic inspections is the equipment used to nondestructively look for cracks. Because of the variability in the shape and locations of cracks, not all cracks will be found during an inspection. Many studies [1,2,3] have shown that the probability that a crack will be detected generally increases with crack size and a commonly used model for the probability of detection (POD) is the cumulative lognormal distribution.

The sizes of cracks that are found during an inspection are commonly used to assess the general state of a fleet of aircraft [4,5]. The distribution of crack sizes can be used to project the likelihood of future failures of key structural details, which is then used to make decisions regarding how often future inspections should be conducted or whether to ground the aircraft. The problem with using the cracks found during an inspection is that the larger cracks are more likely to be found than the smaller cracks. There is an inherent bias to the sampling procedure when relying on nondestructive inspection systems to find the cracks.

Additional complications result from the fact that the crack size distribution changes in time and the individual aircraft in a fleet are inspected at different times. Although the nominal inspection period may be 500 hours, aircraft are scheduled for inspections based on availability and operational needs. The difference in inspection times must also be considered when using crack sizes from routine maintenance inspections.

Two models are developed in this paper to account for the impact of using inspection results to estimate flaw size distributions in aging aircraft. The first provides the basic structure for accommodating the bias that results from using NDE to find cracks. In the first model, all inspections are assumed to have occurred at the same time. The second method incorporates a stochastic crack growth model that is commonly used to make projections of aircraft safety to account for inspections made at different flight hours. The last part of this paper discusses methods for improving the model in future research.

II. Analysis For A Single Inspection Time

A common method for developing estimates of parameters of a model for the distribution of a random variable is the method of maximum likelihood [6]. The likelihood is a function of the model parameters and is proportional to the probability density function given by the model. It is referred to as the likelihood because values of the random variable have already been collected so that the density function evaluated at the observed data represent the relative likelihood of different values of the parameters. Maximum likelihood estimates (MLE's) are chosen to maximize the likelihood for the observed data.

In an aging aircraft, each location that is inspected could have a crack, however most of these cracks are too small to be reliably detected. The resulting set of lengths of detected cracks does not, therefore, represent a complete sample. This is referred to as censoring because the cracks that were missed in the inspection should have been part of the sample but were eliminated due to the inadequacies of the inspection system. The likelihood function must be modified to account for the likelihood of censoring, or missing, individual flaws.

In typical reliability studies, censoring is based on a fixed value, or a fixed number of failures. For example, a study of the reliability of light bulbs may test 100 light bulbs for 1000 hours. A few of the bulbs will fail during the 1000 hours and their life time would be known, but the only information about the bulbs that didn't; fail is that their life times are longer than 1000 hours.

In estimating flaw size distributions, censoring could occur at any length, because even large flaws have a small chance of being missed. A typical inspection system bases detection of cracks on the strength of a signal response that results from the interaction of the crack with some form of applied energy. For example, in ultrasonic inspections, an ultrasonic wave is projected into the component and the magnitude of the wave that is reflected from the crack is measured. A positive crack indication results if the response signal exceeds a specified detection threshold. Censoring is therefore based on the response signal which is correlated with the length of the crack This puts a different twist on the standard method of incorporating censoring into the likelihood function.

The assumptions for developing the appropriate likelihood function are:

1) There are a fixed number of inspection locations, say n, that are examined during the programmed maintenance,
2) Inspection location i will have a crack of length a_i which is assumed to be a random variable with density function $f(a;\underline{\theta})$ where $\underline{\theta}$ is a vector of parameters for the flaw size distribution, and
3) the inspection reliability is known to be POD(a) so that the probability that the crack in location i is detected is $POD(a_i)$.

The probability that a random inspection results in a miss is given by:

$$Q^*(\theta) = \int_0^\infty (1 - POD(a)) f(a;\underline{\theta}) da \tag{1}$$

which incorporates the randomness of the flaw size and the proportion of each flaw size that will be missed. The likelihood for $\underline{\theta}$ for a null inspection (that is, no crack indication) is then $Q^*(\underline{\theta})$. The probability density function for the sizes of cracks that are found is:

$$f_D(a) = POD(a) f(a;\underline{\theta}) / (1 - Q^*(\theta)) \tag{2}$$

and the likelihood of $\underline{\theta}$ for a crack that is detected and found to have size a is $f_D(\underline{\theta};a)$. The likelihood function for a complete data set is the product of the likelihoods for the individual inspection results.

In practice, the flaw size distribution is often modeled as a log normal distribution and the cumulative lognormal distribution function is used for the POD function. Using these assumptions the contributions to the likelihood function for the parameters γ and δ become

$$L(\gamma,\delta;a) = \begin{cases} \Phi\left(\dfrac{\mu-\gamma}{\sqrt{\sigma^2+\delta^2}}\right) & \text{if crack is missed} \\ \Phi\left(\dfrac{\ln(a)-\mu}{\sigma}\right)\dfrac{1}{a\delta\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(a)-\gamma}{\delta}\right)^2} & \text{if found} \end{cases} \tag{3}$$

Where μ and σ are the parameters of the POD function with μ equal to the logarithm of the crack length that is detected 50% of the time and σ is a steepness parameter for the POD function.

These likelihood equations were used to estimate the parameters of the distribution of crack sizes that were inspected in a set of 255 inspections in which 104 cracks were found. A plot of the apparent crack sizes versus time of inspection is shown in Figure 1.
Although the inspections were performed at different times, they appear to be a random sample from a lognormal distribution. Estimation of the flaw size distribution parameters was implemented assuming that there was no time difference to illustrate the basic concept of accounting for the random censoring.

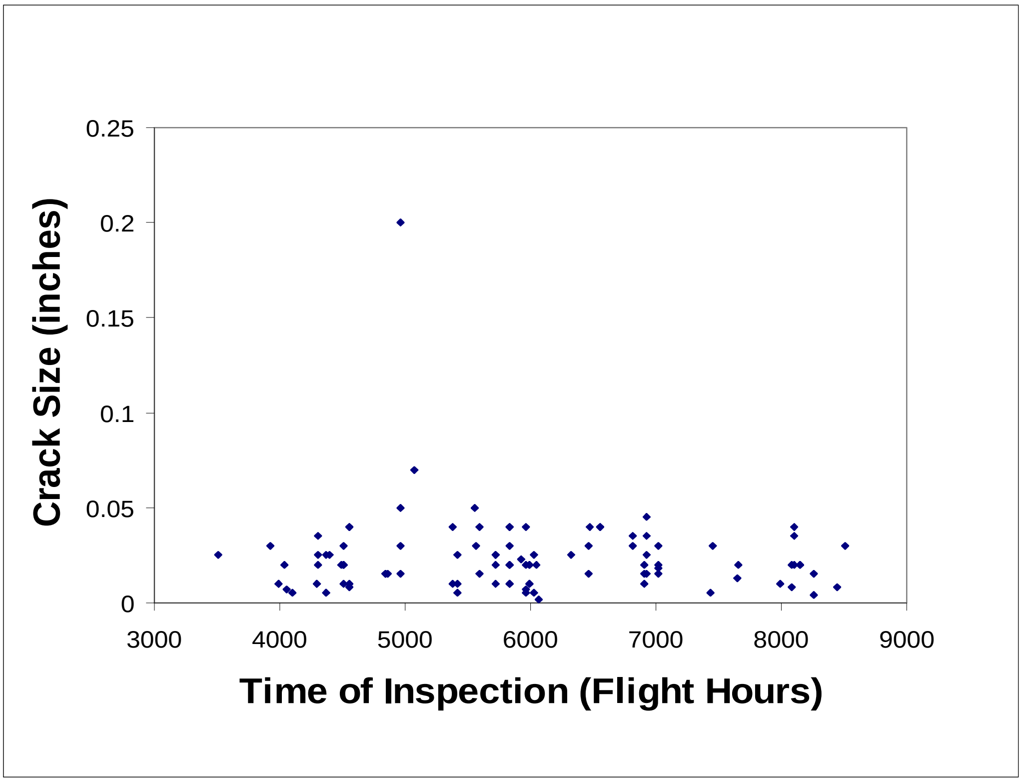

FIGURE 1. Apparent crack sizes of detected cracks.

Figure 2 Shows a plot of the estimated cumulative distribution function of the cracks that were inspected. Note that plot of the maximum likelihood estimate of the distribution function is well to the left of the empirical distribution function of the sizes of the cracks that were detected. Incorporating information about the POD function has corrected for the bias of seeing a larger percentage of big cracks.

Figure 3 shows a similar comparison using density functions instead of cumulative distribution functions. The dashed line shows the estimate of the density function that would result from only using the cracks that were detected, while the solid line shifted to the left shows the impact of accounting for random censoring that is correlated with crack size. The censored estimates put more weight on smaller cracks which are detected less frequently.

III. Analysis For Inspections Performed At Different Times

The impact of the time of the inspection was not accounted for in the first analysis. A second analysis was attempted to account for the change in the flaw size distribution as a function of time. The model that was used to project the change of the flaw size distribution in time is the model used to generate equivalent initial flaw sizes and is often used in risk analyses of aging aircraft. The concept is that percentiles of the flaw size distribution follow the nominal crack grow curve for the structural detail in time. Figure 4 shows a schematic of the concept.

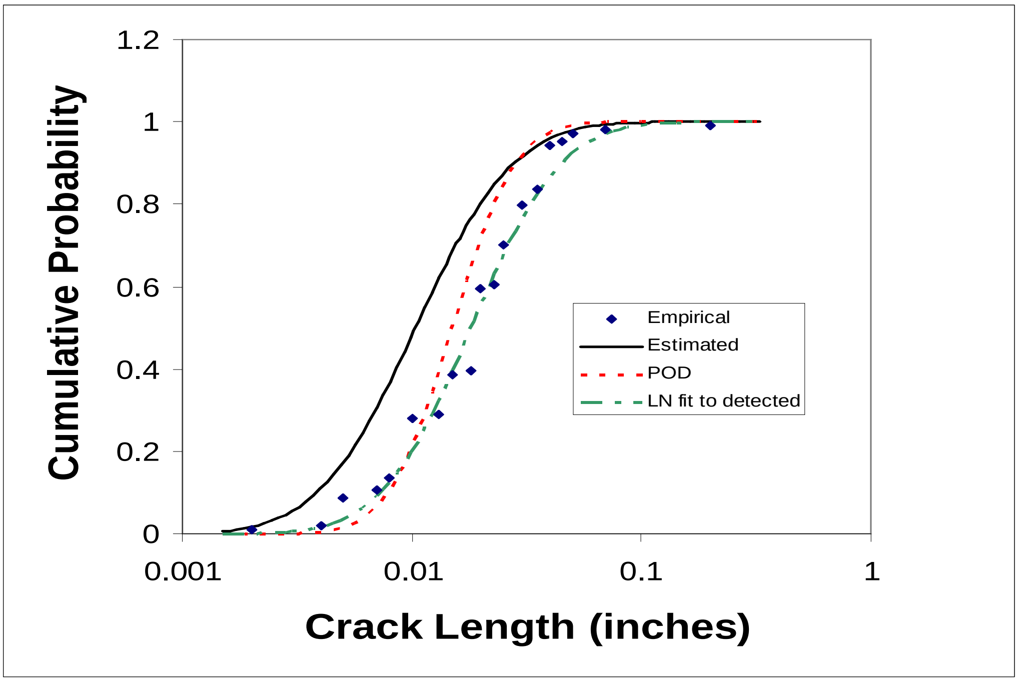

FIGURE 2. Comparison of the estimated crack size distribution function with the apparent distribution of detected cracks.

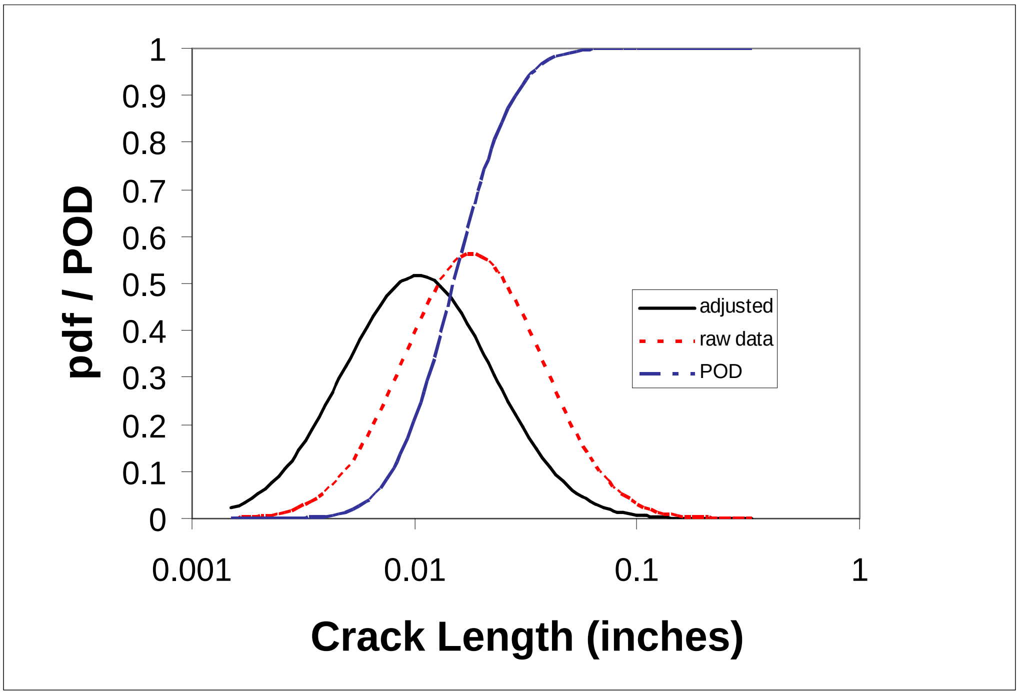

FIGURE 3. Comparison of the estimated density function of the crack size distribution with the distribution of detected cracks.

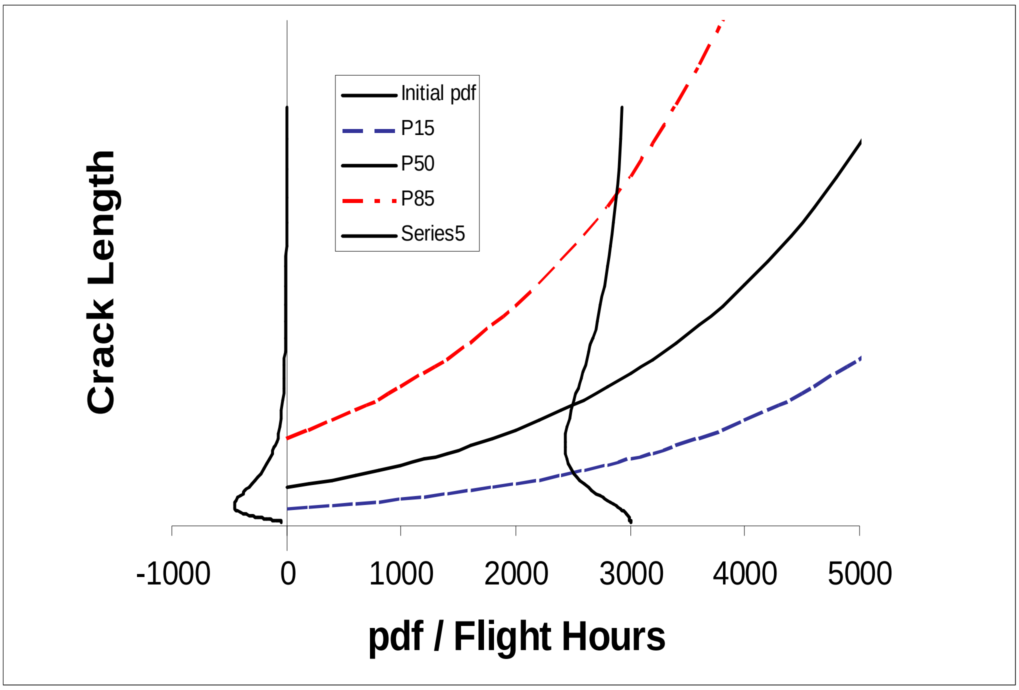

FIGURE 4. Schematic of the stochastic crack growth model.

The impact of information about the crack growth model is that the flaw size distribution has an additional parameter to specify the time of the inspection. This parameter is known, so it does not have to be estimated, however it does impact the likelihood function. The contributions to the likelihood equations become:

$$Q^*(\theta;t) = \int_0^\infty (1 - POD(a)) f(a;\underline{\theta},t)da$$

for cracks that are missed and

$$L(\theta;a) = POD(a) f(a;\underline{\theta},t)$$

for cracks that are found.

The nominal crack growth curve for the structural detail that was used in this study is a close fit to an exponential curve. The equivalent initial flaw size model for an exponential crack growth curve is

$$a(t) = a_0 e^{\lambda t}$$

where a_0 is the crack length at time 0 and λ is the exponential growth rate.

With the assumption of the log normal distribution for the initial flaw size, and an exponential crack growth curve, the distribution of flaws at time t is also an exponential distribution with mean and standard deviation of logarithm of crack length of $\gamma + \lambda$*t and δ where γ is the mean of the logarithm of crack size at time 0.

The specific likelihood functions for lognormal crack sizes, exponential crack growth and the lognormal POD function become:

$$L(\gamma,\delta;a) = \begin{cases} \Phi\left(\dfrac{\mu-(\gamma+\lambda t)}{\sqrt{\sigma^2+\delta^2}}\right) & \text{if crack is missed} \\ \Phi\left(\dfrac{\ln(a)-\mu}{\sigma}\right)\dfrac{1}{a\delta\sqrt{2\pi}}e^{-\frac{2}{2}\left(\frac{\ln(a)-(\gamma+\lambda t)}{\delta}\right)^2} & \text{if found} \end{cases} \tag{4}$$

where γ and δ are the mean and standard deviation of the logarithm of crack length for the crack size distribution at time 0. This would typically be the equivalent initial crack size distribution.

The SAS software converged to estimates of γ and δ indicating strong statistical significance, however, a plot of the data indicates problems with the crack growth model. Figure 5 shows the indicated crack sizes versus time of inspection, overlaid with plots of percentiles of the crack size distribution based on the stochastic crack growth model. It is clearly evident that the indicated crack sizes do not follow this model. The statistical test in the SAS routine did not address the adequacy of the model, only whether or not the parameters are equal to 0.

One reason the model did not fit well is that the equivalent initial flaw size (EIFS) concept may not model crack growth in real aircraft usage. The EIFS concept was developed primarily in laboratory tests which are highly controlled. It is not unreasonable that the variability in usage and environment would cause problems with the model.

Another possible problem is that the indicated crack lengths used in this study were derived from the signal response of the NDE system, not from direct measurements of the cracks. It is possible that inspectors tend to calibrate signal response to a constant range of crack lengths, and that the time dependence is lost. In this case, the first analysis is appropriate; however, the crack length distribution would pertain to the indicated crack sizes from the signal response rather than the true crack size distribution.

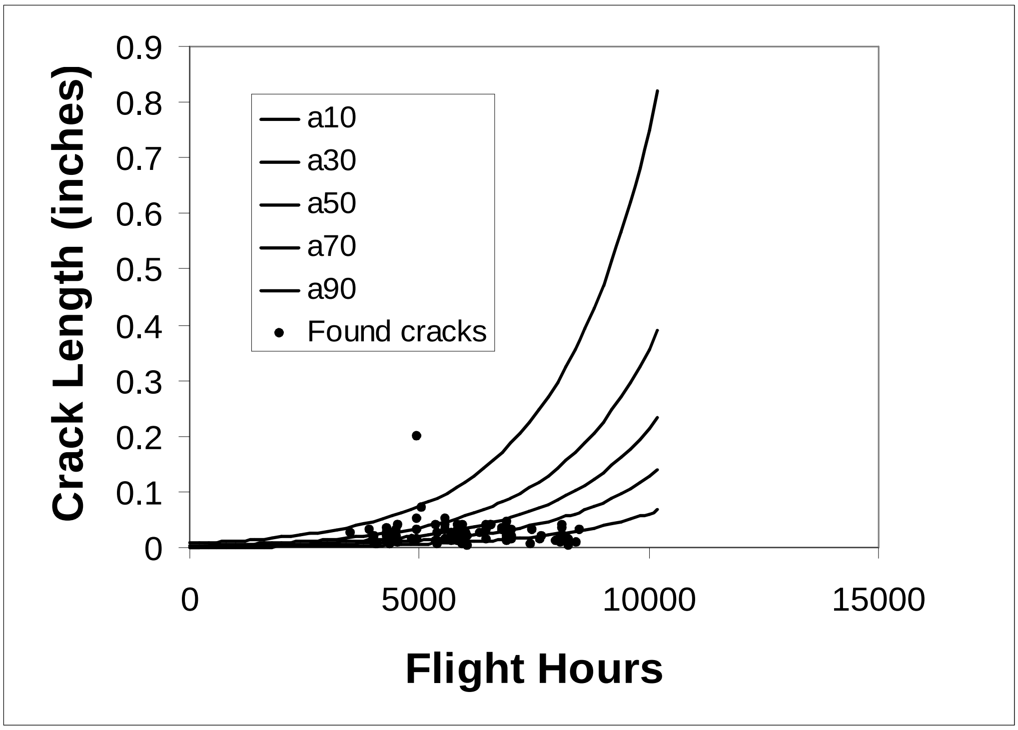

FIGURE 5. Comparison of crack sizes with the stochastic crack growth model

IV. Simulation Study

A simulation study was conducted to demonstrate the potential for the analysis method if accurate flaw size data are available. The exponential crack growth curve model was used along with a lognormal equivalent initial flaw size distribution. The parameter values were chosen to approximate the real data that were collected. The mean and standard deviation of the mean crack size for the initial flaw size distribution were $\gamma = -7.5$, $\delta = 0.7$ which correspond to a median crack size of 0.00055 inches and a 90th percentile crack length of 0.0017 inches. The exponential crack growth rate parameter was $\lambda = 0.000463$ and the POD parameters were $\mu = -3.5$ and $\sigma = 0.5$, which correspond to a crack length that is detected 50% of the time of 0.03 inches and a crack length that is detected 90% of the time of 0.069 inches.

Figure 6 shows a summary of the simulated analysis. Inspections were simulated at various times in the life of the aircraft to approximate realistic aircraft experience. The fitted crack growth model is plotted on top of the simulated crack lengths. The cracks that were missed are plotted as red squares and the cracks that were found are plotted as black diamonds. The fitted crack growth percentiles show a good fit to the simulated crack growth data.

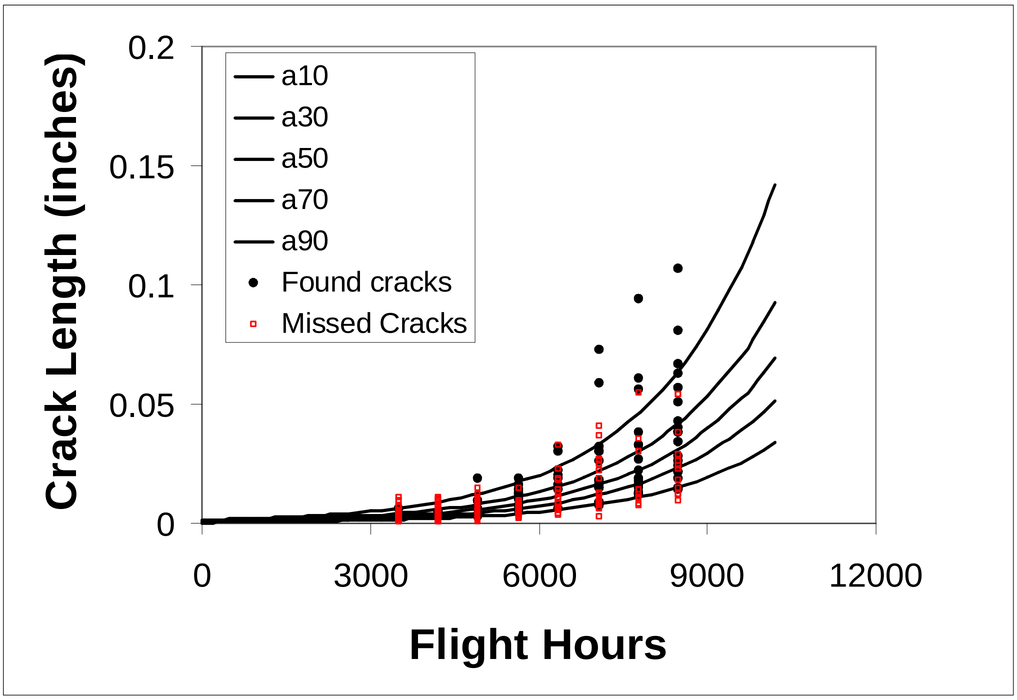

FIGURE 6. Plot of the simulated data and the model fit for.

V. Problems with Estimating POD from Inspection Results

The methods discussed above assume that the POD function is known and although extensive studies have been conducted to estimate the POD function for a wide range of inspections, there are concerns that these studies do not accurately reflect the capability of field inspections. There are many factors that impact the ability of NDE systems to detect cracks and many of them cannot be reproduced in designed evaluation studies. It is impractical to add known flawed structures to field inspections because of the possibility that they are not properly accounted for and end up in service. It is desirable to estimate the POD function directly from routine maintenance inspections.

In theory, it is a simple matter of making the parameters of the POD function unknown values in the likelihood function and maximizing with respect to both the crack size distribution and POD function parameters. The equations are somewhat straightforward to set up and the SAS system was used to try to estimate both sets of parameters with the data used above, however the algorithms did not converge. The problem is that the POD function is a conditional probability given the crack sizes are known. When the true crack sizes are not known there is not enough information to uniquely estimate both the crack size and POD function parameters.

Figure 7 illustrates the problem with convergence by plotting the likelihood function versus pairs of parameters. The upper left plot labeled LKD, is a plot of the likelihood versus the two POD function parameters with the crack length distribution parameters set to their true values. The figure has a shelf appearance, which is the source of the convergence problem in SAS. This surface is nearly flat in the POD parameters which leads to a wide range of values that are suitable maximum likelihood estimates. The other three plots show the likelihood versus: the scale parameters, μ and γ, for the POD function and crack size distribution (LKDs), the location parameters, σ and δ, for the POD function and crack size distribution (LKDm), and the parameters for the crack size distribution (LKDp).

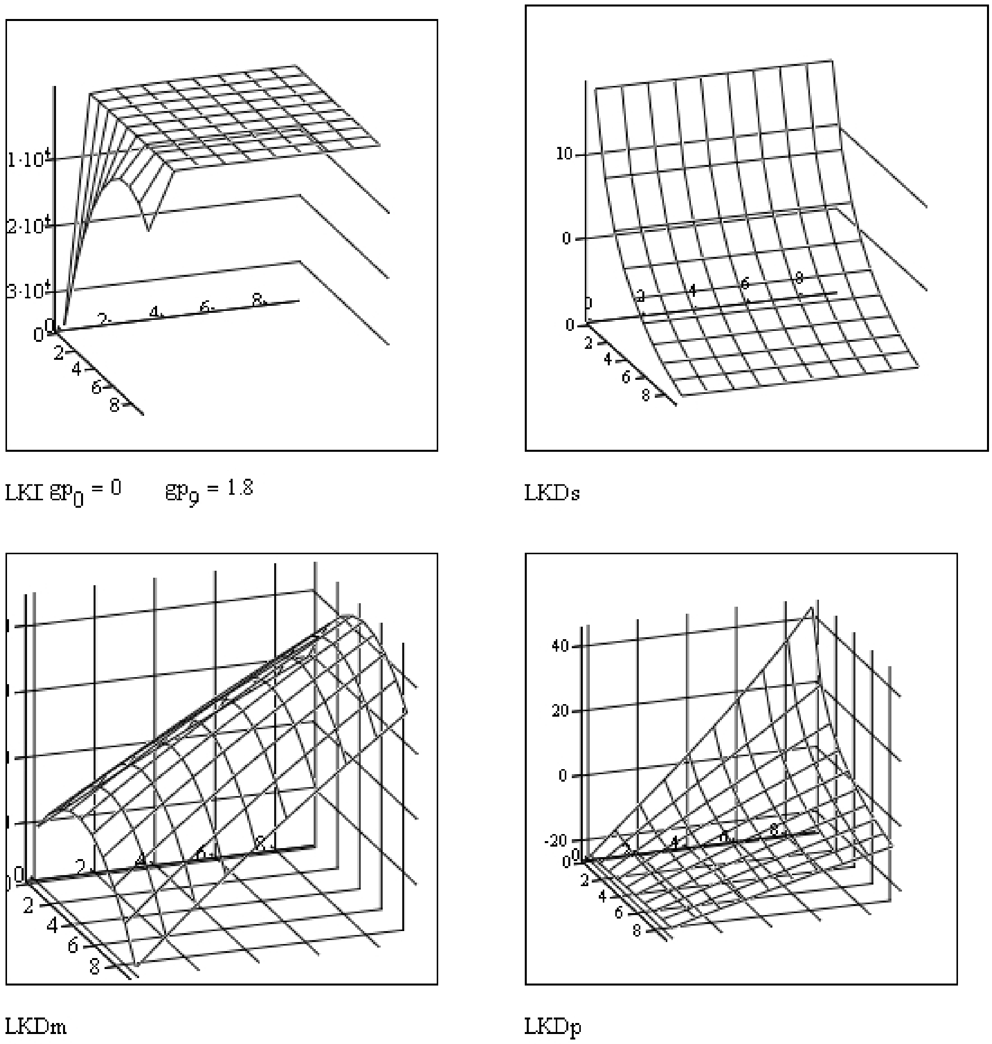

FIGURE 7. Likelihood surface for estimating both crack size distribution and POD function parameters from inspection results.

Current research is looking into other estimation criteria including method of moments estimation, least squares estimation and Bayesian estimation.

VI. Recommendations And Conclusions

This study has shown the need to account for the selection bias induced by using NDE to find cracks when estimating the distribution of cracks that are in the aircraft. An attempt was made to incorporate a commonly used stochastic crack growth model to accommodate inspections performed at different points in the life of the aircraft. Problems with the manner in which flaw sizes were determined were identified that could invalidate the estimates that incorporate a crack growth model to account for the time of inspection.

A simulation study demonstrated the ability of the technique when accurate flaw sizes are available for detected cracks. Future work should place an emphasis on collecting inspection results with crack lengths that have been verified by independent means.

The possibility of estimating both the POD function parameters and he crack size distribution parameters was investigated. The method of maximum likelihood proved to be inadequate to uniquely estimate the parameters. Other methods are being considered in current research programs.

VII. Acknowledgements

Funding was provided by the NDE Branch of the Air Force Research Laboratory. The SAS® Statistical Analysis Software was used to compute maximum likelihood estimates. Microsoft Excel® was used to create the graphs and Mathcad® was used to generate the simulated data.

VIII. References

1. Hovey, P. W. and A. P. Berens, "Evaluating Crack Detection Capability for Aircraft Risk Analysis," The 1997 Proceedings of the Section on Physical and Engineering Sciences of the American Statistical Association, Washington, D. C., 1998.
2. Berens, A.P. and Hovey, P.W., "The Sample Size and Flaw Size Effects in NDI Reliability Experiments," Review of Progress in Quantitative Nondestructive Evaluation 4, Plenum Press, NY, 1985.
3. Berens, A. P. and Hovey, P. W. "Statistical Methods for Estimating Crack Detection Probabilities," Probabilistic Fracture Mechanics and Fatigue Methods: Application for Structural Design and Maintenance, ASTM STP 798, American Society for Testing and Materials, pp. 7994, 1983.
4. Berens, A., Hovey, P., and Miedlar, P., "Aging aircraft maintenance planning," Safety and Reliability of Engineering Systems and Structures, Proceedings of the 9th International Conference on Structural Safety and Reliability, ICOSSAR'05, Rome, Italy, June 19 – 23, 2005, G.Augusti, G. Schueller, M. Ciampoli, editors, Millpress, Rotterdam Netherlands.
5. Berens, A. P. and Hovey, P. W., "Quantifying NDI Capability for Damage Tolerance Analyses", Review of Progress in Quantitative Nondestructive Evaluation, Volume 3, Plenum Press, New York, 1984.
6. Meeker, William Q. and Louis A. Escobar, Statistical Methods For Reliability Data, Wiley, New York, 1998.